Water transport on graphs

Olle Häggström [*]   Timo Hirscher [†]

Chalmers University of Technology

April 16, 2015

Abstract

If the nodes of a graph are considered to be identical barrels – featuring different water levels – and the edges to be (locked) water-filled pipes in between the barrels, one might consider the optimization problem of how much the water level in a fixed barrel can be raised with no pumps available, i.e. by opening and closing the locks in an elaborate succession. This problem originated from the analysis of an opinion formation process and proved to be not only sufficiently intricate in order to be of independent interest, but also algorithmically complex. We deal with both finite and infinite graphs as well as deterministic and random initial water levels and find that the infinite line graph, due to its leanness, behaves much more like a finite graph in this respect.

1   Introduction

Imagine a plane on which rainwater is collected in identical rain barrels, some of which are connected through pipes (that are already water-filled). All the pipes feature locks that are normally closed. If a lock is opened, the contents of the two barrels which are connected via this pipe start to level, see Figure 1. If one waits long enough, the water levels in the two barrels will be exactly the same, namely lie at the average $\frac{a+b}{2}$ of the two water levels (a and b) before the pipe was unlocked.

After a rainy night in which all of the barrels accumulated a certain amount of precipitation we might be interested in maximizing the water level in one fixed barrel by opening and closing some of the locks in carefully chosen order.

In order to mathematically model the setting, consider an undirected graph $G = (V, E)$, which is either finite or infinite with bounded maximal degree. Furthermore, we can assume without loss of generality that G is connected
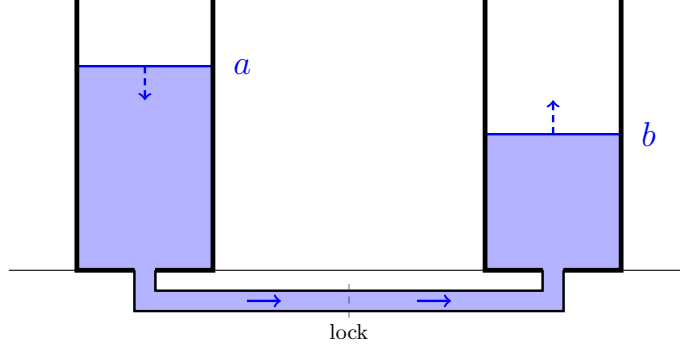
1

Figure 1: Levelling water stages after just having opened a lock.

and simple, that means having neither loops nor multiple edges. Every vertex is understood to represent one of the barrels and the pipes correspond to the edges in the graph. The barrels themselves are considered to be identical, having a fixed capacity $C > 0$.

Given some initial profile $\{\eta_0(u)\}_{u \in V} \in [0, C]^V$, the system is considered to evolve in discrete time and in each round we can open one of the locked pipes and transport water from the fuller barrel into the emptier one. If we stop early, the two levels might not have completely balanced out giving rise to the following update rule for the water profile: If in round k the pipe $e = \langle x, y \rangle$ connecting the two barrels at sites x and y, with levels $\eta_{k-1}(x) = a$ and $\eta_{k-1}(y) = b$ respectively, is opened and closed after a certain period of time, we get

$$\begin{array}{rcl} \eta_k(x) & = & a + \mu_k\,(b - a) \\ \eta_k(y) & = & b + \mu_k\,(a - b) \end{array} \tag{1}$$

for some $\mu_k \in [0, \frac{1}{2}]$, which we assume can be chosen freely by appropriate choice of how long the pipe is left open. All other levels stay unchanged, i.e. $\eta_k(w) = \eta_{k-1}(w)$ for all $w \in V \setminus \{x, y\}$.

The quantity of interest is then defined as follows:

Definition 1
For a graph $G = (V, E)$, an initial water profile $\{\eta_0(u)\}_{u \in V}$ and a fixed vertex $v \in V$ (the *target vertex*), let a *move sequence* be given by a list of edges and time spans that determines which pipes are opened (in chronological order) and for how long. Let then $\kappa(v)$ be defined as the supremum over all water levels that are achievable at v with move sequences consisting of finitely many rounds, i.e.

$$\kappa(v) := \sup\{r \in \mathbb{R}, \text{ there exists } T \in \mathbb{N}_0 \text{ and a move sequence s.t. } \eta_T(v) = r\}.$$

Readers familiar with mathematical models for social interaction processes might note that (1) basically looks like the update rule in the opinion formation

2

process given by the so-called *Deffuant model* for consensus formation in social networks (as described in the introduction of [5]), only μ can change from update to update and the bounded confidence restriction is omitted. This however is no coincidence: The situation described in the context above arises naturally in the analysis of the Deffuant model where the question is how extreme an opinion can a fixed agent possibly get given an initial opinion profile on a specified network graph, if the interactions take place appropriately.

In order to tackle this question, Häggström [4] invented a non-random pairwise averaging procedure, which he proposed to call *Sharing a drink* (SAD). This procedure – which is the main focus of the second section – was originally considered on the infinite line graph only, i.e. the graph $G = (V, E)$ with $V = \mathbb{Z}$ and $E = \{\langle v, v + 1 \rangle, \ v \in \mathbb{Z}\}$, but can immediately be generalized to any graph (see Definition 2) and is dual to the water transport described above in a sense to be made precise in Lemma 2.1.

In Section 3, we will deal with the water transport problem on finite graphs. After formally introducing the idea of optimal move sequences, we investigate both their essential building blocks and the effect of simultaneously opened pipes. In subsection 3.3, being a collection of examples, we will in fact deal with both situations – the one in which we consider the initial water levels to be deterministic and the other in which they are random. In the latter case $\kappa(v)$ obviously becomes a random variable as well since it strongly depends on the initial profile. On non-transitive graphs (see Definition 8) its distribution can moreover depend on the chosen vertex v – even for i.i.d. initial water levels, see Example 3.2.

In the fourth section, we extend the complexity consideration touched upon in some of the examples from Section 3. We show that it is an NP-hard problem to determine $\kappa(v)$ for a given finite graph, target vertex v and initial water profile in general, something that might be considered as a valid excuse for the fact that we are unable to give a neat general solution when it comes to optimal move sequences in the water transport problem on finite graphs, as dealt with in Section 3.

As opposed to the two precedent sections, Section 5 is devoted to infinite graphs. We consider i.i.d. initial water levels (with a non-degenerate marginal distribution) and detect a remarkable change of behavior: On the infinite line graph, the highest achievable water level at a fixed vertex depends on the initial profile in the sense that is has a non-degenerate distribution, just like on any finite graph. If the infinite graph contains a neighbor-rich half-line (see Definition 7), however, this dependence becomes degenerate: For any vertex $v \in V$, the value $\kappa(v)$ almost surely equals the essential supremum of the marginal distribution. This fact makes the infinite line graph quite unique: It constitutes the only exception among all infinite quasi-transitive graphs, to the effect that $\kappa(v)$ is a non-degenerate random variable – an observation which is captured in the last theorem: the nonetheless central Theorem 5.3.

2 Connection to the SAD-procedure

Let us first repeat the formal definition of the SAD-procedure:

Definition 2
For a graph $G = (V, E)$ and some fixed vertex $v \in V$, define $\{\xi_0(u)\}_{u \in V}$ by setting

$$\xi_0(u) = \begin{cases} 1 & \text{for } u = v \\ 0 & \text{for } u \neq v. \end{cases}$$

In each time step, an edge $\langle x, y \rangle$ is chosen and the profile $\{\xi_0(u)\}_{u \in V}$ updated according to the rule (1) with $\{\xi_k(u)\}_{u \in V}$ in place of $\{\eta_k(u)\}_{u \in V}$. One can interpret this procedure as a full glass of water initially placed at vertex v (all other glasses being empty), which is then repeatedly shared among neighboring vertices by each time step choosing a pair of neighbors and pouring a μ_k-fraction of the difference from the glass containing more water into the one containing less. Let us refer to this interaction process as *Sharing a drink (SAD)*.

Just as in [4], the SAD-procedure can be used to describe the composition of the contents in the water barrels after finitely many rounds of opening and closing pipe locks. The following lemma corresponds to La. 3.1 in [4], but since the two dual processes (water transport and SAD) evolve in discrete time in our setting, the proof simplifies somewhat.

Lemma 2.1
Consider an initial profile of water levels $\{\eta_0(u)\}_{u \in V}$ on a graph $G = (V, E)$ and fix a vertex $v \in V$. For $T \in \mathbb{N}_0$ define the SAD-procedure that starts with $\xi_0(u) = \delta_v(u)$ (see Definition 2) and is dual to the chosen move sequence in the water transport problem in the following sense: If in round $k \in \{1, \ldots, T\}$ the water profile is updated according to (1), the update in the SAD-profile at time $T - k \in \{0, \ldots, T-1\}$ takes place along the same edge and with the same choice of μ_k. Then we get

$$\eta_T(v) = \sum_{u \in V} \xi_T(u) \, \eta_0(u). \tag{2}$$

PROOF: We prove the statement by induction on T. For $T = 0$, the statement is trivial and there is nothing to show. For the induction step fix $T \in \mathbb{N}$ and assume the first pipe opened to be $e = \langle x, y \rangle$. According to (1) we get

$$\eta_1(u) = \begin{cases} \eta_0(u) & \text{if } u \notin \{x, y\} \\ (1 - \mu_1) \, \eta_0(x) + \mu_1 \, \eta_0(y) & \text{if } u = x \\ (1 - \mu_1) \, \eta_0(y) + \mu_1 \, \eta_0(x) & \text{if } u = y. \end{cases}$$

Let us consider $\{\eta_1(u)\}_{u\in V}$ as some initial profile $\{\eta'_0(u)\}_{u\in V}$. By induction hypothesis we get

$$\eta'_{T-1}(v) = \sum_{u\in V} \xi'_{T-1}(u)\,\eta'_0(u)$$

$$= \sum_{u\in V\setminus\{x,y\}} \xi'_{T-1}(u)\,\eta_0(u) + \Big((1-\mu_1)\,\xi'_{T-1}(x) + \mu_1\,\xi'_{T-1}(y)\Big)\,\eta_0(x)$$

$$+ \Big((1-\mu_1)\,\xi'_{T-1}(y) + \mu_1\,\xi'_{T-1}(x)\Big)\,\eta_0(y),$$

where $\eta'_{T-1}(v) = \eta_T(v)$ and $\{\xi'_k(u)\}_{u\in V}$, $0 \le t \le T-1$, is the SAD-procedure corresponding to the move sequence after round 1. As by definition the SAD-procedure ξ arises from ξ' by adding an update at time $T-1$ along edge e with parameter μ_1, we find $\xi_k(u) = \xi'_k(u)$ for all $k \in \{0,\dots,T-1\}$ and $u \in V$ as well as

$$\xi_T(u) = \begin{cases} \xi_{T-1}(u) = \xi'_{T-1}(u) & \text{if } u \notin \{x,y\} \\ (1-\mu_1)\,\xi_{T-1}(x) + \mu_1\,\xi_{T-1}(y) & \text{if } u = x \\ (1-\mu_1)\,\xi_{T-1}(y) + \mu_1\,\xi_{T-1}(x) & \text{if } u = y, \end{cases}$$

which establishes the claim. $\qquad\qquad\square$

In the following sections, we want to consider not only deterministic but also random initial profiles of water levels. Having this mindset already, it might be useful to halt for a moment and realize that the statement of Lemma 2.1 deals with a deterministic duality that does not involve any randomness (once the initial profile and the move sequence are fixed).

Before we turn to the task of rising water levels, let us prepare two more auxiliary results. The first one follows directly from the energy argument that was used in the proof of Thm. 2.3 in [4]:

Lemma 2.2
Given an initial profile of water levels $\{\eta_0(u)\}_{u\in V}$ on a graph $G = (V,E)$, fix a finite set $A \subseteq V$ and a set $E_A \subseteq E$ of edges inside A that connects A. If we open the pipes in E_A – and no others – in repetitive sweeps for times long enough such that $\mu_k \ge \varepsilon$ for some fixed $\varepsilon > 0$ in each round (cf. (1)), then the water levels inside the set A approach a balanced average, i.e. converge to the value $\frac{1}{|A|}\sum_{v\in A}\eta_0(v)$. The corresponding dual SAD-profiles started with $\xi_0(u) = \delta_v(u)$, $u \in V$, converge uniformly to $\frac{1}{|A|}\delta_A$ for all $v \in A$.

PROOF: Let us define the energy after round k inside A by

$$W_k(A) = \sum_{v\in A} \big(\eta_k(v)\big)^2.$$

A short calculation reveals that an update of the form (1) reduces the energy by $2\mu_k^2\,(b-a)^2$, where the updated water levels were a and b respectively. If μ_k is bounded away from 0, the fact that $W_k(A) \ge 0$ for all k entails that the

difference in water levels $|b - a|$ before a pipe is opened can be larger than any fixed positive value only finitely many times. In effect, since any pipe in E_A is opened repetitively we must have $|\eta_k(u) - \eta_k(v)| \to 0$ as $k \to \infty$ for all edges $\langle u, v \rangle \in E_A$. As the updates are average preserving, the first part of the claim follows from the fact that E_A connects A.

The second part of the lemma follows by applying the same argument to the dual SAD-procedure. □

The following lemma constitutes an extremely narrowed variant of Thm. 2.3 in [4] which applies to graphs more complex than line graphs as well and will come in useful in Example 3.3:

Lemma 2.3
Fix a (connected) graph $G = (V, E)$ and a vertex $v \in V$. For any $w \in V \setminus \{v\}$, the supremum of $\xi_k(w)$ taken over all times k and SAD-procedures started with $\xi_0(u) = \delta_v(u)$, $u \in V$, is less than or equal to $\frac{1}{2}$.

PROOF: If the SAD-procedure is started with a full glass of water at $v \neq w$, the assumption that the amount at w can rise above $\frac{1}{2}$ leads to the following contradiction: Assume k to be the first time s.t. $\xi_k(w) > \frac{1}{2}$. Then in round k node w necessarily pooled the water with some neighbor u, that had more water than w. But since this relation is preserved by an update, it implies

$$\xi_k(w) + \xi_k(u) \geq 2\,\xi_k(w) > 1,$$

which is impossible as the amount of water shared always sums to 1. □

To round off these preliminary considerations, let us collect some results about SAD-profiles from [4] – partly already mentioned – into a single lemma for convenience.

Lemma 2.4
Consider the SAD-procedure on a line graph, started in vertex v, i.e. with $\xi_0(u) = \delta_v(u)$, $u \in V$.

(a) The SAD-profiles achievable on line graphs are all unimodal.

(b) If the vertex v only shares the water to one side, it will remain a mode of the SAD-profile.

(c) The supremum over all achievable SAD-profiles started with δ_v at another vertex w equals $\frac{1}{d+1}$, where d is the graph distance between v and w.

The results in [4] actually all deal with the infinite line graph, but it is evident how the arguments used immediately transfer to finite line graphs. Part (a) hereby corresponds to La. 2.2 in [4], part (b) to La. 2.1 and part (c) to Thm. 2.3. The argument Häggström [4] used to prove the statement in (c) for the infinite line graph can in fact be generalized to prove the result for trees without much effort, as was done by Shang (see Prop. 6 in [7]).

In fact, we believe that not only the cut back statement from Lemma 2.3 but also the natural generalization of Thm. 2.3 in [4] holds true for general graphs. Our attempts to prove the generalization to non-tree graphs have, however, turned out unsuccessful.

3 Water transport on finite graphs

In this section, we consider the underlying network to be finite, i.e. $|V| = n \in \mathbb{N}$. In order to increase the water level at our fixed site v one could in principle start by greedily trying to connect the barrels with the highest water levels to the one at v. However, optimizing this idea is far from being trivial. Let us first define optimal move sequences and then reveal some properties and building blocks that they share.

Definition 3
For fixed $v \in V$ and a given initial water profile $\{\eta_0(u)\}_{u \in V}$ let $\varphi \in (E \times [0, \frac{1}{2}])^T$, where $\varphi_k = (e_k, \mu_k)$, be called a *finite move sequence* if $T \in \mathbb{N}_0$. φ is a *finite optimal move sequence* if opening the pipes $e_1, \ldots, e_T$ in chronological order, each for the period of time that corresponds to μ_k in (1), will lead to the final value $\eta_T(v) = \kappa(v)$.

For any move sequence $\varphi \in (E \times [0, \frac{1}{2}])^T$, we will denote by $\{\xi_T(u)\}_{u \in V}$ the SAD-profile that corresponds to φ via the duality laid down in Lemma 2.1.

If no finite optimal move sequence exists, let us call $\Phi = \{\varphi^{(m)}, \ m \in \mathbb{N}\}$ an *infinite type optimal move sequence*, provided that $\varphi^{(m)} \in (E \times [0, \frac{1}{2}])^{T_m}$ is a finite move sequence for each $m \in \mathbb{N}$, achieving $\eta_{T_m}(v) > \kappa(v) - \frac{1}{m}$ and the SAD-profiles $\{\xi_{T_m}(u)\}_{u \in V}$ dual to $\varphi^{(m)}$ converge pointwise to a limit $\{\xi(u)\}_{u \in V}$ as $m \to \infty$.

It is tempting to assume that in the case where no finite optimal move sequence exists, we could get away with an infinite move sequence instead of a sequence of finite move sequences Φ as described above. However this is not the case, see Example 3.6.

Lemma 3.1
Take the network $G = (V, E)$ to be finite, and fix the target vertex v as well as the initial water profile. Then the existence of an optimal move sequence is guaranteed and the following simplification will not change its performance: In an optimal move sequence, without loss of generality we can assume $\mu_k = \frac{1}{2}$ for all k.

PROOF: By the very definition of $\kappa(v)$, the existence of optimal move sequences (however not necessarily finite ones) is guaranteed: Let $A \subseteq [0, 1]^V$ denote the set of achievable SAD-profiles. Its closure $\overline{A}$ in $([0, 1]^V, \| \cdot \|_2)$ is bounded and therefore compact. Given the initial water profile $\{\eta_0(u)\}_{u \in V}$, the function

$$f := \begin{cases} [0, 1]^V \to [0, C] \\ \{\xi(u)\}_{u \in V} \mapsto \sum_{u \in V} \xi(u) \, \eta_0(u) \end{cases}$$

7

is continuous. Hence there exists a closed subset F of $\overline{A}$ on which f achieves its maximum $\kappa(v)$ over $\overline{A}$. The SAD-profiles dual to finite optimal move sequences are given by $F \cap A$. If $F \cap A = \emptyset$ and $\Phi = \{\varphi^{(m)}, \ m \in \mathbb{N}\}$ is a collection of finite move sequences s.t. $\varphi^{(m)} \in (E \times [0, \frac{1}{2}])^{T_m}$ and $\eta_{T_m}(v) > \kappa(v) - \frac{1}{m}$ for all $m \in \mathbb{N}$, we can assume without loss of generality that the corresponding SAD-profiles $\{\xi_{T_m}(u)\}_{u \in V}$ have a limit $\{\xi(u)\}_{u \in V}$ (by passing on to a subsequence if necessary) as $\overline{A}$ is compact. This turns Φ into an infinite type optimal move sequence and the limit of its dual SAD-profiles necessarily lies in F.

Assume now that the first move in a sequence $\varphi \in (E \times [0, \frac{1}{2}])^T$ is to open the lock on pipe $e_1 = \langle x, y \rangle$ for a time corresponding to $\mu_1 \in [0, \frac{1}{2}]$ in (1).

Without loss of generality we can assume $\eta_0(x) \geq \eta_0(y)$ (which in turn implies $\eta_1(x) \geq \eta_1(y)$). If we look at the SAD-profile $\{\xi'_{T-1}(u)\}_{u \in V}$ corresponding to $\varphi' := (\varphi_2, \dots, \varphi_T) \in (E \times [0, \frac{1}{2}])^{T-1}$ – in effect we look at the outcome of the move sequence after the first step applied to the new initial water profile $\{\eta_1(u)\}_{u \in V}$ – we can distinguish two cases: either $\xi'_{T-1}(x) \geq \xi'_{T-1}(y)$ or $\xi'_{T-1}(x) < \xi'_{T-1}(y)$. In the first case changing μ_1 to 0, i.e. erasing the first move will not decrease the water level finally achieved at v, see (2). In the second case the same holds for changing μ_1 to $\frac{1}{2}$. Since we can consider any step in the move sequence to be the first one applied to the intermediate water profile achieved so far, this establishes the claim for finite optimal move sequences.

As any finite move sequence can be simplified in this way without worsening its outcome, the argument applies to the elements of a sequence $\Phi = \{\varphi^{(m)}, \ m \in \mathbb{N}\}$ of finite move sequences and thus to infinite type optimal move sequences as well. $\qquad\square$

3.1 Macro moves

When it comes to the opening and closing of pipes, it is not self-evident how far things change if we allow pipes to be opened simultaneously. First of all one has to properly extend the model laid down in (1) by specifying how the water levels behave when more than two barrels are connected at the same time. In order to keep things simple, let us assume that the pipes are all short enough and of sufficient diameter such that we can neglect all kinds of flow effects. Moreover, let us take the dynamics to be as crude as can be by assuming that the water levels of the involved barrels approach their common average in a linear and proportional fashion, which is made more precise in the following definition.

Definition 4
Given a graph $G = (V, E)$, let $A \subseteq V$ be a set of at least 3 nodes and $E_A \subseteq E$ a set of edges inside A that connects A. A *macro move* on E_A (or simply A) will denote the action of opening all pipes that correspond to edges in E_A in some round k simultaneously and will – analogously to (1) – change the water levels for all vertices $u \in A$ to

$$\eta_k(u) = (1 - 2\mu_k)\, \eta_{k-1}(u) + 2\mu_k\, \overline{\eta}_{k-1}(A), \quad \text{where } \overline{\eta}_{k-1}(A) = \frac{1}{|A|} \sum_{w \in A} \eta_{k-1}(w)$$

is the average over the set A after round $k-1$ and $\mu_k \in [0, \frac{1}{2}]$.

First of all, Lemma 2.1 transfers immediately and almost verbatim to move sequences including macro moves: In a move sequence with a macro move on the set A in the first round, we get the water levels

$$\eta_1(u) = \begin{cases} \eta_0(u) & \text{if } u \notin A \\ (1 - 2\mu_1)\, \eta_0(u) + 2\mu_1\, \overline{\eta}_0(A) & \text{if } u \in A. \end{cases}$$

If $\{\xi_{T-1}(u),\ u \in V\}$ and $\{\xi_T(u),\ u \in V\}$ are such that

$$\eta_T(v) = \sum_{u \in V} \xi_T(u)\, \eta_0(u) = \sum_{u \in V} \xi_{T-1}(u)\, \eta_1(u),$$

we find by comparing the coefficient of $\eta_0(u)$

$$\xi_T(u) = \begin{cases} \xi_{T-1}(u) & \text{if } u \notin A \\ (1 - 2\mu_1)\, \xi_{T-1}(u) + \sum_{w \in A} 2\mu_1 \frac{\xi_{T-1}(w)}{|A|} & \text{if } u \in A, \end{cases}$$

which is the SAD-profile originating from the very same macro move applied to $\{\xi_{T-1}(u),\ u \in V\}$. With this tool in hand, we can prove the following extension of Lemma 3.1:

Lemma 3.2

Take the network $G = (V, E)$ to be finite, and fix the target vertex v as well as the initial water profile.

(a) *Even if we allow macro moves, the statement of Lemma 3.1 still holds true, i.e. reducing the range of μ_k from $[0, \frac{1}{2}]$ to $\{0, \frac{1}{2}\}$ in each round k does not worsen the outcome of optimal move sequences.*

(b) *The sharp upper bounds on achievable water levels are not changed if we allow for pipes to be opened simultaneously. In other words, the supremum $\kappa(v)$ of water levels achievable at a vertex v, as characterized in Definition 1, stays unchanged if we allow move sequences to include macro moves.*

PROOF:

(a) Just as in Lemma 3.1, we consider a move sequence consisting of finitely many (macro) moves – say again $T \in \mathbb{N}$ – and especially the SAD-profile dual to the moves after round 1, denoted by $\{\xi_{T-1}(u),\ u \in V\}$. If the first action is a macro move on the set A, let us divide its nodes into two subsets according to whether their initial water level is above or below the initial average across A:

$$A_a := \{u \in A,\ \eta_0(u) \geq \overline{\eta}_0(A)\} \quad \text{and} \quad A_b := \{u \in A,\ \eta_0(u) < \overline{\eta}_0(A)\}.$$

If $\sum_{u \in A_a} \xi_{T-1}(u) \leq \sum_{u \in A_b} \xi_{T-1}(u)$, changing μ_1 to $\frac{1}{2}$ will not decrease the final water level achieved at v. If instead $\sum_{u \in A_a} \xi_{T-1}(u) \geq \sum_{u \in A_b} \xi_{T-1}(u)$, the same holds for erasing the first move (i.e. setting $\mu_1 = 0$).

(b) Obviously, allowing for pipes to be opened simultaneously can if anything increase the maximal water level achievable at v. However, any such macro move can be at least approximated by opening pipes one after another. Levelling out the water profile on a set of more than 2 vertices completely will correspond to the limit of infinitely many single pipe moves on the edges between them (in a sensible order).

Let us consider a finite move sequence φ including macro moves on the sets $A_1 \ldots, A_l$ (in chronological order). From part (a) we know that with regard to the final water level achievable at v we can assume w.l.o.g. that all moves are complete averages (i.e. $\mu_k = \frac{1}{2}$ for all k). Fix $\varepsilon > 0$ and let us define a finite move sequence $\overline{\varphi}$ including no macro moves in the following way: We keep all the rounds in φ in which pipes are opened individually. For the macro move on A_i, $i \in \{1, \ldots, l\}$, we insert a finite number of rounds in which the pipes of an edge set E_{A_i} connecting A_i are opened in repetitive sweeps such that the water level at each vertex $u \in A_i$ is less than $\frac{\varepsilon}{2^i}$ away from the average across A_i after these rounds. Note that Lemma 2.2 guarantees that this is possible.

As opening pipes leads to new water levels being convex combinations of the ones before, the differences of individual water levels caused by replacing the macro moves add up to $\sum_{i=1}^{l} \frac{\varepsilon}{2^i} < \varepsilon$ in the worst case. Consequently, the final water level achieved at v by $\overline{\varphi}$ is at most ε less than the one achieved by φ. Since $\varepsilon > 0$ was arbitrary, this proves the claim.

Note however that the option of macro moves can make a difference when it comes to the attainability of $\kappa(v)$, see Example 3.6. □

Remark
Lemma 3.2 (a) states that even for macro moves, there is nothing to be gained by closing the pipes before the water levels have balanced out completely. A macro move on the edge set E_A with $\mu_k = \frac{1}{2}$ can be seen as the limit of infinitely many single edge moves on E_A in the sense of Lemma 2.2 – a connection that does not exist for macro moves with $\mu_k \in (0, \frac{1}{2})$. We believe that there always exists a finite optimal move sequence if macro moves are allowed. We state this as an open problem.

Due to Lemmas 3.1 and 3.2 we can assume w.l.o.g. that the parameters μ_k in optimal move sequences are always equal to $\frac{1}{2}$ in each round, hence omit them and consider the move sequence to be a list of pipes (i.e. $\varphi \in E^T$) only. We can incorporate an optimal move sequence in which more than one pipe is opened at a time into Definition 3 by either allowing φ_k, for $k \in \{0, \ldots, T\}$, to be a subset of E with more than one element on which the levelling takes place or by viewing φ as a limiting case of move sequences $\{\varphi^{(m)}, m \in \mathbb{N}\}$, in which pipes are opened separately, that form an infinite type optimal move sequence Φ – as just described in the proof of the lemma.

In the sequel however – if not otherwise stated – we will stick to the initial regime where pipes are opened one at a time.

3.2 Optimizing the move sequence

Closely related to the water transport idea is the concept of *greedy lattice animals* as introduced by Cox, Gandolfi, Griffin and Kesten [2]. The vertices of a given graph G are associated with an i.i.d. sequence of non-negative random variables and a greedy lattice animal of size n is then defined to be a connected subset of n vertices containing the target vertex v and maximizing the sum over the associated n random variables. Since we do not care about the size of the lattice animal, let us slightly change this definition:

Definition 5
For a fixed graph $G = (V, E)$, target vertex v and water levels $\{\eta(u)\}_{u \in V}$, let us call $C \subseteq V$ a *lattice animal (LA)* for v if C is connected and contains v. C is a *greedy lattice animal (GLA)* for v if it maximizes the average of water levels over such sets. This average will be considered as its value

$$\text{GLA}(v) := \frac{1}{|C|} \sum_{u \in C} \eta(u).$$

By Lemma 2.2, it is clear that $\text{GLA}(v) \leq \kappa(v)$. In fact, for the majority of settings – consisting of a graph G, a target vertex v and an initial water profile $\{\eta_0(u)\}_{u \in V}$ – strict inequality holds and we can do better than just pooling the amount of water collected in an appropriately chosen connected set of barrels including the one at v.

Furthermore we know from Lemma 3.2 (a) that w.l.o.g. the last move of any finite optimal move sequence will be to pool the amount of water allocated in a connected set of vertices including v. This greedy lattice animal for v in the intermediate water profile created up to that point in time can be more advantageous than the one in the initial water profile if we imply the following improving steps first:

1) Improving bottlenecks
   Let us call a vertex u a *bottleneck* of the GLA C for v if $u \in C \setminus \{v\}$ and $\eta(u) < \text{GLA}(v)$. Clearly, each bottleneck u has to be a cutting vertex for C (otherwise we could just remove it to improve the GLA). If there exists a connected subset of vertices C_u including u which has a higher average water level than $C_u \cap C$, the value of the GLA for v is improved if the water collected in C_u is pooled first (see Figure 2). Note that C_u might involve more vertices from C than just u, see Example 3.5.

2) Enlargement
   The second option to raise the value of the GLA C for v is to apply the idea above to a vertex u in the vertex boundary of C in order for the original GLA to be enlarged to a set of vertices in which u is a bottleneck. For this to be beneficial, there has to exist a connected set of vertices C_u in $V \setminus C$ including u with the following property: The average water level in C_u is smaller than $\text{GLA}(v)$ – otherwise it would be part of C – but is

11

raised above this value after improving the potential bottleneck u using water located in $V \setminus C$ (see Figure 2 below).
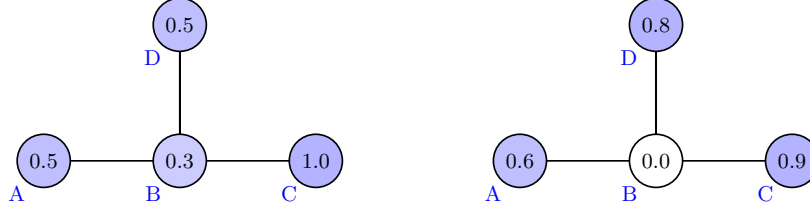

Figure 2: If A is the target vertex, the GLA on the left is $\{A, B, C\}$ (having value 0.6) and the bottleneck B can be improved by first opening the pipe $\langle B, D \rangle$.
The GLA for A with respect to the water profile on the right is $\{A\}$, but can be enlarged to $\{A, B, C\}$ if the potential bottleneck B is improved by opening the pipe $\langle B, D \rangle$ first.

3) Choose optimal chronological order

When applying the improving techniques just described, it is essential to choose the optimal chronological order of doing things. Besides the fact that improving bottlenecks and enlarging the GLA has to be done before the final averaging, situations can arise in which different sets of vertices can improve the same bottleneck or the other way round that more than one bottleneck can be improved using non-disjoint sets of vertices, see the set-ups in Figure 3.
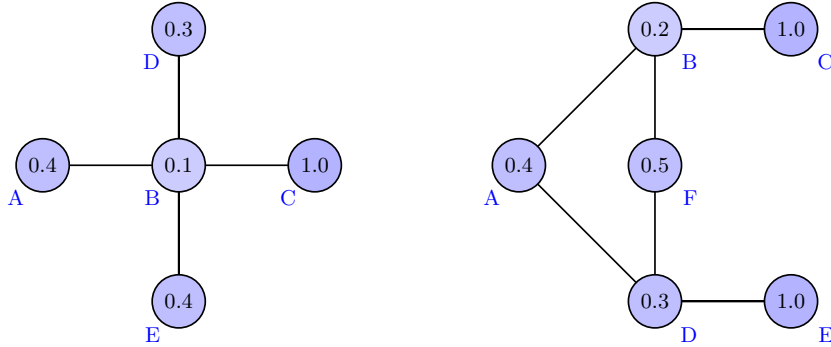

Figure 3: If A is the target vertex, the GLA on the left is $\{A, B, C\}$ (having value 0.5). Improving the bottleneck B can be done using D or E and is most effective if the pipe $\langle B, D \rangle$ is opened first, then $\langle B, E \rangle$.
The GLA for A with respect to the graph on the right is $\{A, B, C, D, E\}$. The water from F can be used to improve both bottlenecks B and D. It is optimal to open pipe $\langle D, F \rangle$ first and then $\langle B, F \rangle$.

Finally, it is worth noticing that lattice animals with lower average than $\mathrm{GLA}(v)$ in the initial water profile sometimes can be improved by the techniques just described to finally outperform the initial GLA and its possible improvements and enlargements (see Example 3.5 and especially Figure 5).

3.3 Examples

Example 3.1
The minimal graph which is non-trivial with respect to water transport is a single edge, in other words the complete graph on two vertices:

$$G = K_2 = (\{1,2\}, \{\langle 1,2 \rangle\}).$$

By the considerations in the previous subsection, we get

$$\kappa(1) = \begin{cases} \eta_0(1) & \text{if } \eta_0(1) \geq \eta_0(2) \\ \frac{\eta_0(1)+\eta_0(2)}{2} & \text{if } \eta_0(1) < \eta_0(2). \end{cases} \tag{3}$$

Let the initial water levels be given by the two random variables U_1 and U_2. From (3) it immediately follows that

$$U_1 \leq \kappa(1) \leq \max\{U_1, U_2\}.$$

If we assume U_1 and U_2 to be independent and uniformly distributed on $[0,1]$, a short calculation reveals the distribution function

$$F_{\kappa(1)}(x) = \begin{cases} \frac{3}{2}x^2 & \text{for } 0 \leq x \leq \frac{1}{2} \\ x - \frac{1}{2}(1-x)^2 & \text{for } \frac{1}{2} \leq x \leq 1, \end{cases}$$

which indeed lies in between $F_{U_1}(x) = x$ and $F_{\max\{U_1,U_2\}}(x) = x^2$, see Figure 4.

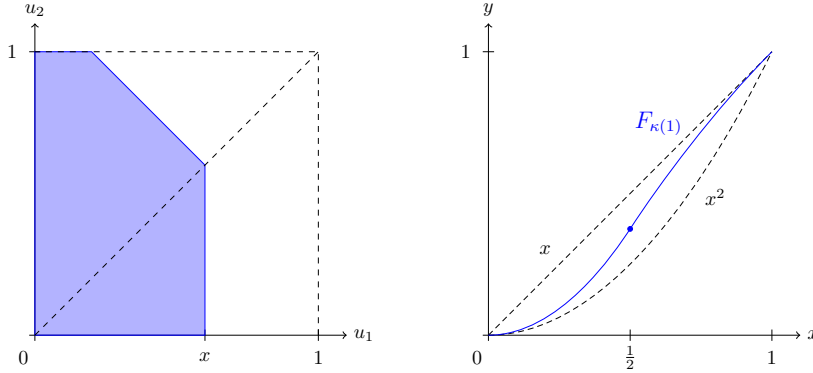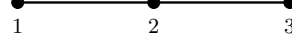Figure 4: On the left a visualization of $\mathbb{P}(\kappa(1) \leq x)$, on the right the distribution function of $\kappa(1)$.

By symmetry, the exact same considerations hold for $\kappa(2)$.

13

Example 3.2
The simplest non-transitive graph (i.e. having
vertices of different kind, see Definition 8) is
the line graph on three vertices:

$$G = (\{1, 2, 3\}, \{\langle 1, 2\rangle, \langle 2, 3\rangle\}).$$

Again by the above considerations, we find the supremum of achievable water
levels at vertex 1 to be

$$\kappa(1) = \max\left\{\eta_0(1), \frac{\eta_0(1)+\eta_0(2)}{2}, \frac{\eta_0(1)+\eta_0(2)+\eta_0(3)}{3}\right\},$$

which is obviously achieved by a properly chosen greedy lattice animal.

Consider the case in which the initial water levels satisfy

$$\eta_0(3) \geq \eta_0(2) \geq \eta_0(1) \quad \text{and} \quad \eta_0(3) > \eta_0(1) \tag{4}$$

Then $\kappa(1) = \frac{\eta_0(1)+\eta_0(2)+\eta_0(3)}{3}$ and there exists no finite optimal move sequence.
This can be seen from the fact that any single move will preserve the inequalities
in (4) and thus we have $\eta_T(1) < \kappa(1) < \eta_T(3)$ for all finite move sequences
$\varphi \in E^T$.

If we consider the initial water levels to be independent and identically
distributed, the (random) supremum of achievable water levels at vertex 2 is
stochastically larger than the one at vertex 1: As $\eta_0(1)$ and $\eta_0(2)$ have the same
distribution so do

$$\kappa(1) \quad \text{and} \quad \max\left\{\eta_0(2), \frac{\eta_0(1)+\eta_0(2)}{2}, \frac{\eta_0(1)+\eta_0(2)+\eta_0(3)}{3}\right\}.$$

The latter is less than or equal to $\kappa(2)$. The maximal value achievable by greedy
lattice animals at vertex 2 is

$$\text{GLA}(2) = \max\left\{\eta_0(2), \frac{\eta_0(1)+\eta_0(2)}{2}, \frac{\eta_0(2)+\eta_0(3)}{2}, \frac{\eta_0(1)+\eta_0(2)+\eta_0(3)}{3}\right\}.$$

The fact that we can average across one pipe at a time and choose the order of
updates allows us to improve over this and gives

$$\kappa(2) = \max\left\{\text{GLA}(2), \frac{1}{2}\left(\eta_0(1) + \frac{\eta_0(2)+\eta_0(3)}{2}\right), \frac{1}{2}\left(\eta_0(3) + \frac{\eta_0(1)+\eta_0(2)}{2}\right)\right\}. \tag{5}$$

To see this, we can take a closer look on the SAD-profiles that can be created
by updates along the two edges $\langle 1, 2\rangle$ and $\langle 2, 3\rangle$ starting from the initial profile
$\xi_0 = (0, 1, 0)$: After one update – depending on the chosen edge – the profile is
given by $\xi_1 = (\frac{1}{2}, \frac{1}{2}, 0)$ or $(0, \frac{1}{2}, \frac{1}{2})$. After the second step we end up with either
$\xi_2 = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ or $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$. All of the corresponding convex combinations appear
in the right hand side of (5). By Lemma 2.2, we know that continuing like this
will finally result in the limiting profile $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. It is not hard to check that any
sequence of two or more updates will lead to a monotonous SAD-profile with a
largest value of at most $\frac{1}{2}$ at one and a smallest value of at least $\frac{1}{4}$ at the other
end. For this reason, it can be written as a convex combination of $(0, 1, 0)$ plus

either $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ and $(0, \frac{1}{2}, \frac{1}{2})$ or $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{2}, 0)$. Consequently, it cannot correspond to a final water level at vertex 2 exceeding the value in (5).

By an elementary calculation, for independent unif$([0, 1])$ initial water levels, we obtain

$$F_{\kappa(1)}(x) = \mathbb{P}(\kappa(1) \leq x) = \int\limits_{s=0}^{x} \int\limits_{t=0}^{\min\{2x-s,1\}} \min\{3x - (s+t), 1\} \, dt \, ds$$

$$= \begin{cases} \frac{8}{3}x^3 & \\ -\frac{11}{6}x^3 + \frac{9}{2}x^2 - \frac{3}{2}x + \frac{1}{6} & \\ -\frac{23}{6}x^3 + \frac{13}{2}x^2 - 2x + \frac{1}{6} & \\ \frac{2}{3}x^3 - \frac{5}{2}x^2 + 4x - \frac{7}{6} & \end{cases} \quad \text{for} \quad \begin{cases} x \in [0, \frac{1}{3}] \\ x \in [\frac{1}{3}, \frac{1}{2}] \\ x \in [\frac{1}{2}, \frac{2}{3}] \\ x \in [\frac{2}{3}, 1] \end{cases},$$

and similarly

$$F_{\kappa(2)}(x) = \int\limits_{s=0}^{x} \int\limits_{t=0}^{\min\{2x-s,1\}} \min\{1, 2x - s, 3x - (s+t), 4x - (2s+t), 2x - \tfrac{s+t}{2}\} \, dt \, ds$$

$$= \begin{cases} 2x^3 & \\ -\frac{14}{3}x^3 + 9x^2 - 4x + \frac{7}{12} & \\ \frac{2}{3}x^3 - 3x^2 + 5x - \frac{5}{3} & \end{cases} \quad \text{for} \quad \begin{cases} x \in [0, \frac{1}{2}] \\ x \in [\frac{1}{2}, \frac{3}{4}] \\ x \in [\frac{3}{4}, 1] \end{cases},$$

which is strictly smaller than $F_{\kappa(1)}(x)$ implying $\kappa(1) \preceq \kappa(2)$, where $\preceq$ denotes the usual stochastic order. Due to the fact that adding the edge $\langle 1, 3 \rangle$ will not give an improvement over the optimal move sequences for vertex 2, this stochastic domination already follows from the fact that $\kappa(1)$ is non-decreasing when adding an edge and the symmetry of K_3.

In fact, when optimizing the move sequence for the middle vertex we can neglect the option of levelling out the profile completely, since for any initial water profile there is a *finite* optimal move sequence $\varphi \in E^T$ achieving

$$\eta_T(2) \geq \tfrac{1}{3} \left(\eta_0(1) + \eta_0(2) + \eta_0(3) \right),$$

as the next example will show.

Example 3.3
Given an initial water profile $\{\eta_0(u)\}_{u \in V}$ and the complete graph K_n as underlying network, we get for any $v \in V$:

$$\kappa(v) = 2^{-l+1} \eta_0(v) + \sum_{i=1}^{l-1} 2^{-i} \eta_0(v_i),$$

where V is ordered such that $\eta_0(v_1) \geq \eta_0(v_2) \geq \cdots \geq \eta_0(v_n)$ with $v = v_l$. Furthermore, this optimal value can be achieved by a finite move sequence.

To see this is not hard having Lemmas 2.1 and 2.3 in mind. If $v = v_1$, the highest water level is already in v and the best strategy is to stay away from

15

the pipes. For $v \neq v_1$, the contribution of vertex v_1 – i.e. the share $\xi_T(v_1)$ in the convex combination of $\{\eta_0(u)\}_{u \in V}$ optimizing $\eta_T(v)$, see (2) – can not be more than $\frac{1}{2}$ by Lemma 2.3. However, this can be achieved by opening the pipe $\langle v, v_1 \rangle$. According to the duality between water transport and SAD, this is what we do last. The argument just used can be iterated for the remaining share of $\frac{1}{2}$ giving that v_2 can contibute at most $\frac{1}{4}$ (given that v_1 contributes most possible) and so on. Obviously, involving vertices holding water levels below $\eta_0(v)$ can not be beneficial, as all vertices are directly connected, so we do not have intermediate vertices being potential bottlenecks.

The optimal move sequence $\varphi \in E^T$, where $T = l - 1$, is then given by

$$\varphi_k = \langle v, v_{l-k} \rangle, \ k = 1, \ldots, l - 1$$

leading to

$$\eta_k(v) = 2^{-k} \eta_0(v) + \sum_{i=1}^{k} 2^{-k+i-1} \eta_0(v_{l-i})$$

and consequently $\eta_T(v) = \eta_{l-1}(v) = \kappa(v)$. Note that the option to open several pipes simultaneously is useless on the complete graph. Furthermore the above move sequence only includes edges to which v is incident, so the very same reasoning holds for the center v of a star graph on n vertices as well.

To determine the optimal achievable value at v we have to sort the n initial water levels first. This can be done using the randomized sorting algorithm 'quicksort' which makes $O(n \log(n)))$ comparisons on average, $O(n^2)$ in the worst case. The calculation of $\kappa(v)$ given the sorted list of initial water levels needs at most $n - 1$ additions and $n - 1$ divisions by 2.

Example 3.4
Expanding Example 3.2, let us reconsider the line graph – this time not on three but n vertices. Let the vertices be labelled 1 through n and let vertex 1 (sitting at one end of the line) be the target vertex. Given an initial water profile $\{\eta_0(i)\}_{i=1}^{n}$, $\kappa(1)$ can be determined by $2n - 2$ arithmetic operations ($n - 1$ additions, $n - 1$ divisions) as it turns out to be

$$\kappa(1) = \max_{1 \leq l \leq n} \frac{1}{l} \sum_{i=1}^{l} \eta_0(i). \tag{6}$$

In other words, $\kappa(1)$ is achieved by averaging over the greedy lattice animal for vertex 1 with respect to the initial water profile (see Definition 5).

In order to establish this, let us first define for any water profile $\{\eta_T(i)\}_{i=1}^{n}$ achievable from $\{\eta_0(i)\}_{i=1}^{n}$ (in finite time $T \in \mathbb{N}_0$) the corresponding normed vector
$$\lambda_T := \tfrac{1}{M} \Big(\eta_T(1), \ldots, \eta_T(n) \Big),$$

which can be understood to be a probability measure on $\{1 \ldots, n\}$ – as (1) preserves the total mass $M := \sum_{i=1}^{n} \eta_0(i)$.

16

Related to this construction, there are two important observations to make: Firstly, if we consider two water profiles on $\{1\ldots,n\}$ with the same total mass M and their corresponding probability measures being such that one stochastically dominates the other, this relation is preserved when the two profiles are exposed to the same update of the form (1), see also La. 2.4 in [4]. Secondly, if $\{\eta_{k+1}(i)\}_{i=1}^n$ arises from $\{\eta_k(i)\}_{i=1}^n$ by an update of the form (1) along an edge $\langle u, u+1 \rangle$, where the water level at $u+1$ is higher than the one at u, the corresponding probability measure λ_{k+1} is stochastically dominated by λ_k.

If we choose to open only this kind of pipes (where an update brings water closer to vertex 1) in a sensible order – in permanent sweeps from $\langle 1, 2 \rangle$ to $\langle n-1, n \rangle$ for example – we get a stochastically decreasing sequence of probability measures $(\lambda_k)_{k \in \mathbb{N}_0}$, whose limit is stochastically dominated by any measure $\frac{1}{M}\big(\eta_T(1), \ldots, \eta_T(n)\big)$, where $T \in \mathbb{N}_0$ and $\{\eta_T(i)\}_{i=1}^n$ was created by updates of the form (1) successively applied to the initial profile $\{\eta_0(i)\}_{i=1}^n$.

Furthermore, following this update scheme the water profiles are tending to a piecewise constant profile: Any relation "the barrel at vertex x holds at most as much water as the one at $x+1$" will be preserved if we only open pipes $\langle u, u+1 \rangle$, where $u+1$ has a water level higher than the one at u. For that reason, already the initial water profile determines two sorts of pipes: If L is minimal in respect of

$$\frac{1}{L} \sum_{i=1}^{L} \eta_0(i) = \max_{1 \leq l \leq n} \frac{1}{l} \sum_{i=1}^{l} \eta_0(i)$$

and $L > u$, the water level at $u+1$ will eventually be higher than the one at u causing infinitely many updates along $\langle u, u+1 \rangle$ (or the very same water level at u and $u+1$) in the sequel. If instead $L \leq u$, either the water level at u is always at least as high as at $u+1$ and $\langle u, u+1 \rangle$ will never be opened or the barrel at u is at some point emptier than the one at $u+1$ leading to infinitely many updates along $\langle u, u+1 \rangle$ or eventually the same water level at u and $u+1$. This establishes the claimed shape of the limiting profile (according to Lemma 2.2). Its optimality can be seen as follows: The probability measure λ_T corresponding to an arbitrary achievable profile $\{\eta_T(i)\}_{i=1}^n$ dominates λ_k for k large enough, hence
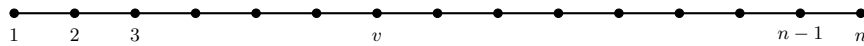
$$\eta_T(1) = M \cdot \lambda_T(1) \leq M \cdot \lambda_k(1) = \eta_k(1) \leq \lim_{k \to \infty} \eta_k(1),$$

and (6) follows.

If we allowed macro moves (opening several pipes simultaneously), the first (and only) move would be to open the pipes $\langle 1, 2 \rangle, \ldots, \langle L-1, L \rangle$.

Example 3.5
Finally, let us consider the line graph on n vertices, with the target vertex v not sitting at one end.

17

Given the initial water levels $\{\eta_0(u), \ 1 \leq u \leq n\}$, let us consider the final SAD-profile $\{\xi(u)\}_{1 \leq u \leq n}$ corresponding to an optimal move sequence (if the move sequence is of infinite type, it is the limit of its dual SAD-profiles we are talking about, cf. Lemmas 2.1 and 3.1).

First of all, from Lemma 2.4 (a) we know that any achievable SAD-profile on line graphs is unimodal (which therefore holds for a pointwise limit of SAD-profiles as well). Let us denote the leftmost maximizer of $\{\xi(u)\}_{1 \leq u \leq n}$ by q and set

$$l := \min\{1 \leq u \leq n, \ \xi(u) > 0\} \quad \text{and} \quad r := \max\{1 \leq u \leq n, \ \xi(u) > 0\}.$$

By symmetry, we can assume without loss of generality $l \leq v \leq q \leq r$ – if $q < v$, the set-up is merely mirrored. Furthermore, let us pick the optimal move sequence such that $\{\xi(u)\}_{1 \leq u \leq n}$ minimizes the distance $q - v$.

The contribution from the nodes $\{q, q+1, \ldots, n\}$ can be seen as a scaled-down version of the problem treated in the previous example: This time the drink to be shared does not amount to 1 but to $\sum_{q \leq u \leq r} \xi(u)$ instead. From Example 3.4 we can therefore conclude that a flat SAD-profile i.e.

$$\xi(q) = \xi(q+1) = \ldots = \xi(r) \tag{7}$$

is optimal. The same holds for the contribution coming from $\{1, 2, \ldots, v-1\}$, i.e.

$$\xi(l) = \xi(l+1) = \ldots = \xi(v-1). \tag{8}$$

In addition to that, from Lemma 2.4 (c) we know $\xi(r) \leq \frac{1}{r-v+1}$.

If $l = v$, part (b) of Lemma 2.4 in turn implies $v = q$. The SAD-profile then features only one non-zero value (namely $\frac{1}{r-v+1}$) and corresponds to the greedy lattice animal for v consisting of the vertices $v, v+1, \ldots, r$. If instead $l < v$ – compared to the balanced average across $\{v, v+1, \ldots, r\}$ just described – the contribution to the final water level at v (cf. (2)) given by

$$\sum_{u=l}^{v-1} \xi(u) \, \eta_0(u) \quad \text{replaced the contribution} \quad \sum_{u=v}^{r} \left(\tfrac{1}{r-v+1} - \xi(u)\right) \eta_0(u), \tag{9}$$

where necessarily $\sum_{u=l}^{v-1} \xi(u) = \sum_{u=v}^{r} \left(\frac{1}{r-v+1} - \xi(u)\right) =: M$. As q is a mode and due to (7) we have

$$\tfrac{1}{r-v+1} - \xi(v) \geq \ldots \geq \tfrac{1}{r-v+1} - \xi(q) = \ldots = \tfrac{1}{r-v+1} - \xi(r). \tag{10}$$

The aforementioned replacement is most beneficial if the weighted average to the right in (9) is made as *small* as possible, keeping M fixed. In view of (10) we can conclude, applying again the ideas from the foregoing example – this time think of the initial profile $C - \eta_0(u)$ considered for $v \leq u \leq r$ only – that this is achieved once more by a balanced average. Hence $l < v$ implies $v < q$ and the just mentioned balanced average has to stretch to the right as far as $q - 1$, i.e. $\xi(v) = \ldots = \xi(q-1) < \frac{1}{r-v+1} = \xi(q)$, since otherwise q would

18

not be the leftmost mode. From this and (8) we find $M = (v - l) \cdot \xi(l) = (q - v) \cdot \left(\frac{1}{r-v+1} - \xi(v) \right)$.

The assumption that $q - v$ was minimized when picking the optimal move sequence considered, forces

$$\sum_{u=l}^{v-1} \xi(u)\, \eta_0(u) \;>\; \sum_{u=v}^{q-1} \left(\tfrac{1}{r-v+1} - \xi(u) \right) \eta_0(u),$$

since otherwise the balanced average across $\{v, v+1, \ldots, r\}$ would have been at least as good. Connecting v to barrels to the left consequently yields an improvement of the final water level at v (in comparison to $\frac{1}{r-v+1} \sum_{u=v}^{r} \eta_0(u)$) to the amount of

$$\sum_{u=l}^{v-1} \xi(u)\, \eta_0(u) - \sum_{u=v}^{q-1} \left(\tfrac{1}{r-v+1} - \xi(u) \right) \eta_0(u) = M \cdot \left(\tfrac{1}{v-l} \sum_{u=l}^{v-1} \eta_0(u) - \tfrac{1}{q-v} \sum_{u=v}^{q-1} \eta_0(u) \right).$$

As a consequence, M must be as large as possible for an optimal move sequence, which means $\xi(l) = \xi(q-1)$ and makes $\{\xi(u)\}_{1 \le u \le n}$ a piecewise constant profile taking on two non-zero values, $\xi(l)$ and $\xi(r)$, as depicted below.
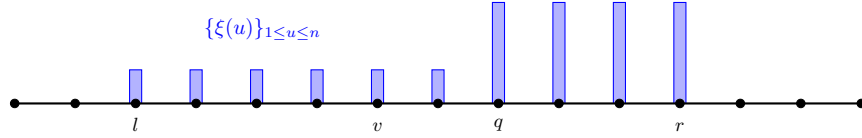
Note that the value $\xi(r) = \frac{1}{r-v+1}$ (and so even $\xi(l)$) is already determined by the choice of l, q and r. In Figure 5 below, a set of initial water levels on the line graph comprising 15 nodes is shown, for which the SAD-profile corresponding to an optimal move sequence is the one shown above. Furthermore, it can be seen from this instance that the GLA with respect to the initial water profile and its possible enhancements can be outperformed by improving another lattice animal as mentioned at the end of Subsection 3.2.
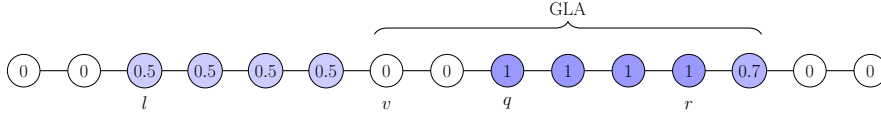

Figure 5: Even for a graph as simple as the line graph, the initial GLA sometimes has little to do with the optimal move sequence.

When it comes to the complexity of finding $\kappa(v)$, we can greedily test all choices for l, q, r – of which there are less than n^3. For each choice at most $n+3$ additions/subtractions and four multiplications/divisions have to be made to

calculate either

$$
\frac{q-v}{(q-l)\,(r-v+1)} \sum_{u=l}^{q-1} \eta_0(u) + \frac{1}{r-v+1} \sum_{u=q}^{r} \eta_0(u) \quad \text{or}
$$

$$
\frac{1}{v-l+1} \sum_{u=l}^{\hat{q}} \eta_0(u) + \frac{v-\hat{q}}{(r-\hat{q})\,(v-l+1)} \sum_{u=\hat{q}+1}^{r} \eta_0(u),
$$

(11)

depending on whether $v \leq q$ or $\hat{q} \leq v$, where $\hat{q}$ is the rightmost mode of $\{\xi(u)\}_{1 \leq u \leq n}$. Even if there might exist SAD-profiles with $q < v < \hat{q}$ corresponding to optimal move sequences, by the above we know that there has to be one with either $v \leq q$ or $\hat{q} \leq v$ as well. The maximal value among those calculated in (11) equals $\kappa(v)$, so the complexity is $\mathcal{O}(n^4)$.

Example 3.6

The preceding example can serve to give a concrete instance in which even an infinite sequence of single edge moves can not achieve the supremum as mentioned after Definition 3.

Consider the line graph on four vertices, the target vertex not to be one of the end vertices and initial water levels as depicted to the right.
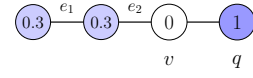
From Example 3.5 we know that the optimal SAD-profile will allocate $\frac{1}{6}$ of the shared glass of water to each of the vertices to the left of v and v itself, the maximal amount of $\frac{1}{2}$ to the rightmost vertex q – showing that $\kappa(v) = 0.6$: First, recall that any SAD-profile on a line graph is unimodal. If q is not the (only) mode, the contribution of v and q has an average of at most 0.5 and thus the SAD-profile in question yields a water level at v of at most 0.5 – see (2). If q is the mode, the SAD-profile is non-decreasing from left to right and thus a flat profile on the vertices other than q uniquely optimal. Finally, to achieve the optimum, the contribution of q has to be maximal, i.e. $\frac{1}{2}$ (see Lemma 2.3).

From the considerations in Thm. 2.3 in [4] it is clear that this SAD-profile, more precisely the value $\frac{1}{2}$ at q, can only be established if the first move is v sharing the drink with q (which corresponds to the last move in the water transport – see Lemma 2.1). Once v starts to share the drink to the left, any other interaction with q will decrease the contribution of the latter and thus put a water level of 0.6 at v out of reach.

To get a flat profile on three vertices, we need however infinitely many single-edge moves (here on e_1 and e_2). An infinite type optimal move sequence is for example given by

$$
\Phi = \{\varphi^{(m)}, \, m \in \mathbb{N}\}, \quad \text{where} \quad \varphi^{(m)} \in E^{T_m}, \, T_m = 2m+1 \quad \text{and}
$$

$$
\varphi^{(m)} = (\underbrace{e_1, e_2, e_1, e_2, \ldots}_{m \text{ pairs}}, \langle v, q \rangle),
$$

achieving

$$
\lim_{m \to \infty} \eta_{T_m}(v) = 0.6 = \kappa(v),
$$

a value that can not be approached by any stand-alone (finite or) infinite sequence of moves.

If we allow macro moves, however, there is a two-step move sequence achieving the water level 0.6 at v: First we open the pipes e_1 and e_2 simultaneously to pool the water of the vertices other than q and in the second round, we open the pipe $\langle v, q \rangle$.

4   Complexity of the problem

In this section, we want to build on the complexity considerations for the water transport on finite graphs from the examples in Section 3. In fact, we want to show that the task of determining whether $\kappa(v)$ is larger or smaller than a given constant – for a generic set-up, consisting of a graph, target vertex and initial water profile – is an NP-hard problem. This is done by establishing the following theorem:

Theorem 4.1
The NP-complete problem 3-SAT can be polynomially reduced to the decision problem of whether $\kappa(v) > c$ or not, for an appropriately chosen water transport instance and constant c.

Before we deal with the design of an appropriate water transport instance in order to embed the satisfiability problem *3-SAT*, let us provide the definition of Boolean satisfiability problems as well as known facts about their complexity.

Definition 6
Let $X = \{x_1, x_2, \ldots, x_k\}$ denote a set of *Boolean variables*, i.e. taking on logic truth values 'TRUE' (T) and 'FALSE' (F). If x is a variable in X, x and $\overline{x}$ are called *literals* over X. A *truth assignment* for X is a function $t : X \to \{T, F\}$, where $t(x) = T$ means that the variable x is set to 'TRUE' and $t(x) = F$ means that x is set to 'FALSE'. The literal x is true under t if and only if $t(x) = T$, $\overline{x}$ is true under t if and only if $t(x) = F$.

A *clause* C over X is a disjunction of literals and *satisfied* by t if at least one of its literals is true under t. A logic formula F is in *conjunctive normal form (CNF)* if it is the conjunction of (finitely many) clauses. It is called *satisfiable* if there exists a truth assignment t such that all its clauses are satisfied under t.

The standard Boolean satisfiability problem (often denoted by *SAT*) is to decide whether a given formula in CNF is satisfiable or not. If we restrict to the case where all the clauses in the formula consist of at most 3 literals it is called *3-SAT* .

3-SAT was among the first computational problems shown to be NP-complete, a result published in a pioneering article by Cook in 1971, see Thm. 2 in [1].

Let us now turn to the task of embedding 3-SAT into an appropriately designed water transport problem that is in size polynomial in n, the number of clauses of the given 3-SAT problem:

Given the logic formula $F = C_1 \wedge C_2 \wedge \ldots \wedge C_n$ in which each of the clauses C_i consists of at most 3 distinct literals, let us define the comb-like graph depicted

in Figure 6. All the white nodes, plus the target vertex v, represent empty barrels. The other ones that are shaded in blue contain water to the amount specified.
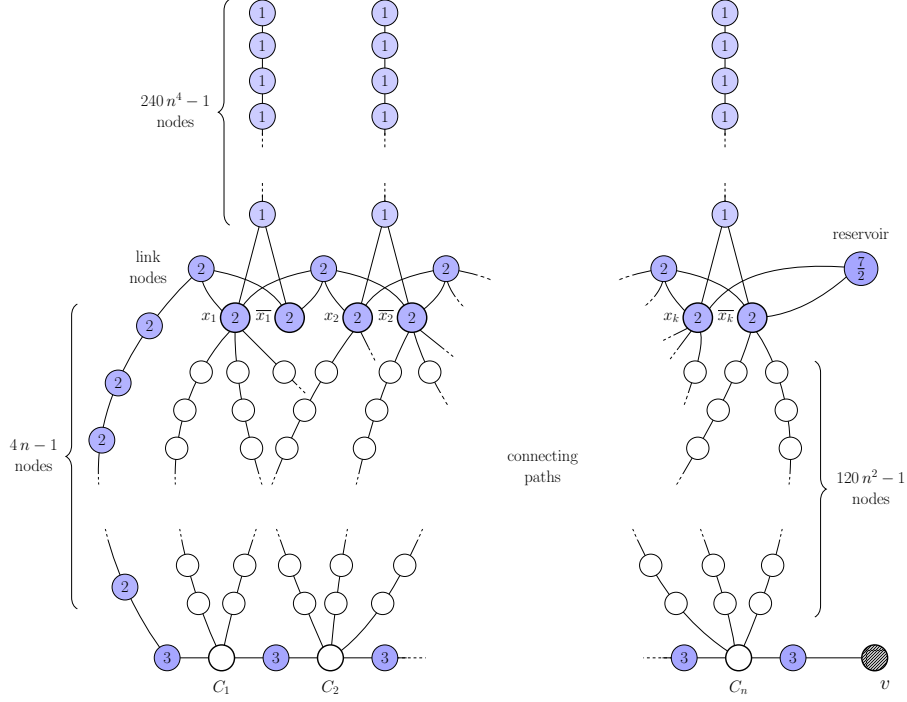

Figure 6: A polynomial reduction of 3-SAT to the water transport problem.

The comb has k teeth, where k is the number of variables appearing in F. Each individual tooth is formed by a line graph on $240\,n^4 - 1$ vertices with water level 1 each. The lower endvertex of the ith tooth is connected to two vertices representing the literals x_i and $\overline{x_i}$, having water level 2 respectively. In between the teeth there are $k-1$ *link nodes*, each of which features itself a water level of 2 and is connected to the four nodes representing literals of consecutive variables – more precisely, the link node in between tooth i and $i+1$ is connected to the vertices $x_i, \overline{x_i}, x_{i+1}, \overline{x_{i+1}}$, for $i \in \{1, \ldots, k-1\}$. The vertices representing $x_k, \overline{x_k}$ are connected to the rightmost link node as well as to an additional vertex featuring a water reservoir of level $\frac{7}{2}$. Left of the first tooth, there is another link node (with water level 2 as well) connected to x_1 and $\overline{x_1}$ as well as by a path to the shaft of the comb, which is described next.

The comb's shaft is made up of a line graph on $2n + 2$ vertices, with the target vertex v to the very right. To the left of v there is a vertex representing a barrel with water level 3 followed by n (empty) barrels that stand for the clauses $C_1, \ldots, C_n$ and are seperated by a vertex with water level 3 respectively. The left endvertex (connected to C_1) features a water level of 3 as well and

is connected to the leftmost link node as mentioned before, namely via a path consisting of $4\,n-1$ nodes with water level 2 each.

Finally, the teeth are connected to the shaft through (disjoint) *connecting paths* from nodes representing literals to nodes representing clauses, where for example $\overline{x_2}$ is linked to C_2 by a path if it appears in this clause. Each of these paths is formed by a line graph on $120\,n^2-1$ vertices representing empty barrels. Note that each clause-node is linked to at most 3 connecting paths, whereas the number of connecting paths originating from a vertex representing a literal can vary between 0 and n.

In connection with the water transport problem originating from a 3-SAT formula F as depicted in Figure 6, we claim the following:

Proposition 4.2
Consider the water transport problem based on the logical formula F, given by the graph, target vertex and initial water profile as depicted in Figure 6.

(a) If F is satisfiable, then the water level at v can be raised to a value strictly larger than 2, i.e. $\kappa(v) > 2$.

(b) If F is not satisfiable, then this is impossible, i.e. $\kappa(v) \leq 2$.

Before we deal with the proof of the proposition, note how it implies the statement of Theorem 4.1: First of all, if F is a 3-SAT formula consisting of n clauses, k cannot exceed $3\,n$. Given this, it is not hard to check that the graph in Figure 6 has no more than $720\,n^5+360\,n^3+9\,n+2$ vertices and maximal degree at most $n+3$ (or 5 if $n=1$). As the initial water levels are all in $\{0,1,2,3,\frac{7}{2}\}$, the size of this water transport instance is clearly polynomial in n. Due to the fact that the value of $\kappa(v)$ can be used to decide whether the given formula F is satisfiable or not – as claimed by Proposition 4.2 – Theorem 4.1 follows.

PROOF OF PROPOSITION 4.2 (a): To prove the first part of the proposition, let us assume that F is satisfiable. Then there exists a truth assignment t with the property that all clauses $C_1, \ldots, C_n$ contain at least one of the k literals that are set true by t. Those can be used to let the water trickle down from the teeth to the line graph at the bottom in an effective way: We assign each clause to one of the true literals under t which it contains. Then, we average the water over k (disjoint) star-shaped trees. Each such tree has a literal $x \in \{x_1, \overline{x_1}, \ldots, x_k, \overline{x_k}\}$ that is true under t as its center and the top node of the tooth above x as well as the nodes representing the clauses assigned to x as leaves (where the clause-nodes are connected to x in the tree via the corresponding connecting paths). If m clauses chose x, there are $240\,n^4 + m \cdot 120\,n^2$ vertices in the tree and the water accumulated amounts to $240\,n^4 + 1$.

By pooling the water along those trees, all the nodes corresponding to clauses can simultaneously be pushed to a water level as close to the average of the corresponding trees as we like (see Lemma 2.2). As $m \leq n$, we can bound these

averages from below by

$$\frac{240\,n^4 + 1}{240\,n^4 + 120\,mn^2} \geq \frac{240\,n^4 + 1}{240\,n^4 + 120\,n^3} > 1 - \tfrac{1}{2n}.$$

So after this procedure, each clause-node will have a water level strictly larger than $1 - \tfrac{1}{2n}$. Note that only one of each pair $\{x_i, \overline{x_i}\}$ was used as a water passage, so there is still a line graph – let us call it *linking path* – consisting of vertices with water level 2 exclusively, from the leftmost link node to the vertex with initial water level $\tfrac{7}{2}$ through all link nodes and the untouched literals (the ones that are false under t).

By another complete averaging – this time over the line graph that consists of the shaft (i.e. the line graph at the bottom in Figure 6), the path to the very left connecting the leftmost link node to the shaft, the linking path just described, as well as the reservoir with level $\tfrac{7}{2}$ at the other end – will push the water level at v beyond

$$\tfrac{1}{6n+2k+2}\left(\tfrac{7}{2} + (2\,k + 4\,n - 1)\cdot 2 + (n+1)\cdot 3 + n\left(1 - \tfrac{1}{2n}\right)\right) = \frac{12\,n + 4\,k + 4}{6\,n + 2\,k + 2} = 2.$$

Consequently, for the case of satisfiable F we verified for the graph depicted in Figure 6: $\kappa(v) > 2$. □

In the proof of the second part of the proposition, we need a rough estimate of how much the water level in a vertex representing a clause can be raised, if only accessed via connecting paths. This is done in the following lemma.

Lemma 4.3
In the comb-like graph depicted in Figure 6, it is impossible to push the water level in a clause-vertex above the value of $1 + \tfrac{1}{2n}$ without opening the pipes to its left or right neighbor.

PROOF: The proof of this claim is a simple comparison with a tree similar to the structure above the node corresponding to some clause C_l. Originating from C_l, there are at most 3 connecting paths that lead to three nodes representing literals. Initially, the node corresponding to C_l and the ones on the connecting paths are empty. Their water level can be raised to almost 1 using water from the teeth of the comb and further using nodes with initial water level higher than 1. The fact that opening pipes always produces convex combinations of the involved water levels (see (2)) guarantees that the total amount of water above a fixed level – cumulated over all barrels – is non-increasing when pipes are opened. Initially, the cumulated amount of water above level 1 in the whole graph is

$$(\tfrac{7}{2} - 1) + 3\,k \cdot (2 - 1) + (4\,n - 1) \cdot (2 - 1) + (n+1) \cdot (3 - 1) \leq 15\,n + \tfrac{7}{2}. \quad (12)$$

For $n \in \mathbb{N}$, this is clearly less than $20n - 1$.

We can mimick any attempt raising the water level at C_l in the comb-graph via its connecting paths in the tree depicted to the right in Figure 7 in such a

way, that the water levels at C_l and on its connecting paths are at any point in time at most as high as the ones in the corresponding part of the comparison tree: If water is routed into the connecting paths above C_l but water levels do
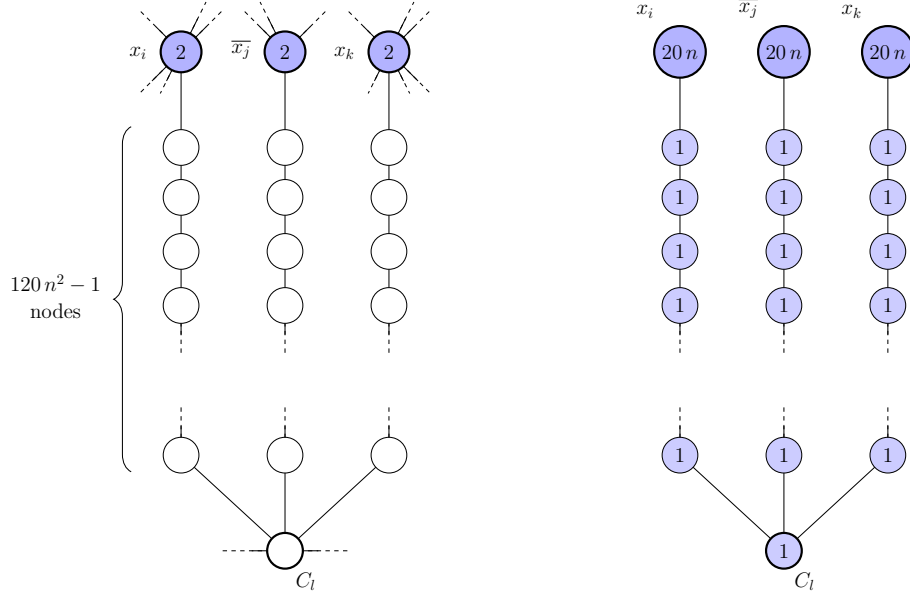

Figure 7: Comparison of the structure above the node representing a clause C_l in the comb-graph with an appropriately tailored tree.

not exceed 1 (e.g. when routing water down from the teeth) we do nothing in the comparison tree. If water from the vertices with initial water level above 1 is introduced into the connecting paths, we introduce the same amount to the corresponding connecting paths in the tree (note that this is possible, as the total amount of water above level 1 in the comb-graph is available in all three leaves of the tree). Every move involving only nodes from the connecting paths depicted and C_l is copied in the tree. This retains the property that the water levels in the tree are not less than the ones in corresponding nodes of the comb-graph and shows that the highest water level achievable at C_l in the tree is an upper bound on the level achievable in the comb-graph. If there are less than 3 connecting paths above C_l in the comb-graph we can either modify the comparison tree accordingly or just not use the extra branches.

By the generalization of Thm. 2.3 in [4] to trees, see the comment after Lemma 2.4, we know that the contribution to the convex combination at C_l from the leaves in the tree is at most 1 divided by the graph distance plus one, i.e. $\frac{1}{120\,n^2+1}$. The water level at C_l in the tree can therefore not exceed

$$1 + 3 \cdot \frac{20\,n}{120\,n^2 + 1} \leq 1 + \frac{1}{2n},$$

which induces the claim. $\qquad\square$

Note that the same argument with only water to the amount of $5\,n-1$ above level 1 available in the leaves would give the upper bound of $1+\frac{15\,n}{120\,n^2}=1+\frac{1}{8n}$, which will be used in the proof of Proposition 4.2 (b) as well.

PROOF OF PROPOSITION 4.2 (b): To check that in case F is not satisfiable we get $\kappa(v)\le 2$ is a bit more involved than the first part: Let us assume the contrary. Then there exists a finite move sequence (involving macro moves say) that achieves a final water level $\eta_T(v)>2$. By the idea in part (a) of Lemma 3.2 we can assume the last move to be the complete average over a connected vertex set A including v. The only barrels with initial water level larger than 2 are the ones left of each clause-node and of v plus the reservoir. Including any node apart from these into the set A, when trying to achieve $\eta_T(v)>2$, can therefore only be beneficial if it is a bottleneck (see the discussion after Definition 5).

Structurally speaking, there are three potential candidates for such a set A:

- a set containing some vertex from a connecting path
- a set containing only vertices from the bottom line graph or
- a set containing the reservoir vertex but no connecting path.

Note that the set we used in the case of satisfiable F was of the third type. We will see in a moment that this is in fact the only relevant candidate for the set A in the sense that the other two do not allow to raise the water level at v above the value of 2, even for a satisfiable formula F.

The first candidate listed is ruled out rather easily: If A contains a vertex from a connecting path, the bottleneck argument forces A to contain the whole corresponding connecting path (recall: a bottleneck has to be a cut vertex between barrels with water levels above average and the target vertex). Then A is of size at least $120\,n^2$ and the amount of water above the water level of 1 is just not sufficient to fill up so many vertices to a level of two: From (12) we know that the water available above level 1 in the whole graph is at most $15\,n+\frac{7}{2}$ initially and non-increasing. The amount in a whole connecting path with water level 2 would be $120\,n^2-1$, so this can definitely not be achieved.

Next, let us assume that A is a subset of the vertices of the bottom line graph – including vertex v and m clauses. Again by the bottleneck argument, we can assume that the leftmost node in A is not a clause-node, i.e. has initial water level 3 (see Figure 8).
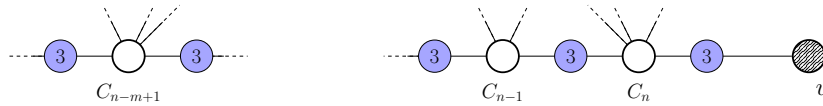

Figure 8: Vertices of the set A considered in the second case.

With the intention to increase the total amount of water inside A before the final averaging, one can try to fill up the clause-nodes. However, from Lemma 4.3 we know that the water level at the clause-nodes (being bottlenecks) in A cannot be pushed much above the level of 1, if accessed via connecting paths only. Further, this makes accessing vertex v through a connecting path and C_n unfavorable. Note that opening the pipes in the bottom line graph in order to connect barrels representing clauses inside A to the ones with water level 2 or 3 outside A might increase the amount of water in A as well, but will raise the water level at the involved clause-nodes to a level that can not be further improved by using links via connecting paths, so it is most beneficial to fill up the clause-nodes with water routed through connecting paths first.

Let us assume that after this first phase, we managed to achieve a water level of $1 + \frac{1}{2n}$ at $C_{n-m+1}, \ldots, C_n$. This might be technically impossible, but surely dominates the water levels achievable using the connecting paths only and simplifies our further considerations. Staying away from the connecting paths, water can only be routed into A via C_{n-m}. If we average over all nodes in A once while doing so, the final averaging is meaningless (because then the effect of any move between this and the last move will be increased if again all pipes inside A are opened). However, since the last move has to involve v we can assume that any move before leaves the pipe on the edge incident to v closed – and thus w.l.o.g. the pipe from C_n towards v as well.

This in turn requires that the connected subset of nodes outside A (incident to the leftmost node in A) that pools its water with a connected subset inside A, including C_{n-m+1} but not v, has an average of at least $2 + \frac{1}{4n}$, as the amount of water inside A would decrease otherwise. In view of Lemma 4.3, the only useful move is therefore to open the pipes along the shaft and through the nodes with initial water level 2 (which they actually might have lost during the first phase) in order to connect the vertex with water reservoir $\frac{7}{2}$ to A. No matter which of the nodes representing literals we include, the water levels in the path connecting the leftmost link node to the shaft and the shaft itself will be dominated by the ones obtained if we pretend that the water above level 2 from the reservoir can be transferred to the leftmost link node without any losses.

Starting with a water level of $\frac{7}{2}$ in the leftmost link node instead, we might increase the amount of water inside A by at most another $\frac{3}{2} \cdot \frac{2m}{2m+4n} \leq \frac{1}{2}$ (as the path has to involve at least $4n$ nodes outside A, the subset inside A is of size at most $2m \leq 2n$ and already has an average of at least 2). Along a line graph, the contribution of the water level from an endvertex to the ones formed as convex combinations along the line is decreasing with the graph distance (see Lemma 2.4 (b)).

Despite our greatest efforts, the total amount of water in A will consequently not exceed the value $3(m+1) + m(1 + \frac{1}{2n}) + \frac{1}{2} \leq 4(m+1)$. Since A consists of $2m + 2$ vertices, leveling out across this set will possibly raise the amount of water in the barrel at v to the level of 2, but not beyond.

Finally, consider A to contain all of the shaft as well as the vertex with water reservoir $\frac{7}{2}$, but no vertex from a connecting path. Then A, being connected,

has to contain the path that consists of $4\,n - 1$ vertices, connecting the shaft to the leftmost link node, as well as a linking path through link nodes and vertices representing literals as described above. However, this time – with F being not satisfiable – it is impossible to fill up all the clause-nodes to a level of about 1, leaving at least one path between the reservoir and the leftmost link node unaffected: In order to reach all clause-nodes, we have to use both x_i and $\overline{x_i}$ as water passage for at least one $i \in \{1, \ldots, k\}$.

In comparison to the case of satisfiable F, we will lose an amount of at least $1 - \frac{1}{2n}$ for each clause-node that is not reached before the final averaging – but likewise an amount of at least 1 in the linking path for each pair $\{x_i, \overline{x_i}\}$ in which both nodes were used as water passage – since their water level of 2 reduces to something less than 1 when water from the tooth above is routed through the node all the way down to a clause-node. By the same token as in Lemma 4.3, the clause-nodes can be filled up to a level of at most $1 + \frac{1}{8n}$ through the connecting paths, as the water available outside A above level 1 is k (from the literals not part of the linking path) plus 1 from a vertex representing a literal on the linking path if we need to route through such (and $k + 1 < 5\,n - 1$). Note that moving water from inside A through a connecting path to a clause will in fact reduce the amount of water in A. Consequently, the set A (which is the same as the one chosen for satisfiable F) still contains $6n + 2k + 2$ vertices, but the amount of water we can allocate in A is at most

$$
\begin{aligned}
& \tfrac{7}{2} + (2\,k + 4\,n - 1) \cdot 2 + (n + 1) \cdot 3 + n\left(1 + \tfrac{1}{8n}\right) - \left(1 - \tfrac{1}{2n}\right) \\
=\ & 12\,n + 4\,k + \tfrac{29}{8} + \tfrac{1}{2n} \\
<\ & 12\,n + 4\,k + 4,
\end{aligned}
$$

for $n \geq 2$. Thus, even in this manner we can not raise the water level at vertex v to a level of 2 or above if F is not satisfiable which contradicts the above assumption and in consequence verifies $\kappa(v) \leq 2$ for this case. $\qquad\square$

As already mentioned, this shows that solving the decision problem "$\kappa(v) > 2$ or $\kappa(v) \leq 2$" for the comb-like graph depicted in Figure 6 solves the corresponding 3-SAT problem as well. Since 3-SAT is an NP-complete problem, we hereby established that any problem in NP can be polynomially reduced to a decision problem minor to the computation of $\kappa(v)$ in a suitable water transport instance – showing that computing $\kappa(v)$ in general is indeed an NP-hard problem.

5 On infinite graphs

This last section is devoted to the water transport problem on infinite graphs. We consider an infinite, connected, simple graph $G = (V, E)$ with bounded maximal degree. The initial water levels $\{\eta_0(u)\}_{u \in V}$ are considered to be i.i.d. with a (non-degenerate) common marginal distribution concentrated on $[0, C]$, for some $C > 0$. The supremum $\kappa(v)$ of achievable water levels at a fixed target vertex $v \in V$ depends on the initial water levels of course, which makes it a random variable as well.

When the vertices of an infinite graph are assigned individual values, the most natural definition of a *global average* across the graph is to look at a fixed sequence of nested subsets of the vertex set, with the property that every vertex is included eventually, and then consider the limit of averages across those subsets (if it exists).

Given i.i.d. initial water levels, the strong law of large numbers tells us that the randomness of the global average – which is non-degenerate on finite graphs – becomes degenerate if we consider infinite graphs, where it will a.s. equal the expectation of the marginal distribution. $\kappa(v)$ however shows a slightly different behavior: In order to determine whether the supremum of achievable water levels at a given vertex v is a.s. constant or not, we have to investigate the global structure of the infinite graph a bit more closely.

If the graph contains a half-line with sufficiently many extra vertices attached to it, the distribution of $\kappa(v)$ becomes degenerate for all $v \in V$ – as stated in Theorem 5.1 and the final remark: One can in fact, with probability 1, push the water level at v to the essential supremum of the marginal distribution. The infinite line graph however is too lean to feature such a substructure and behaves therefore much more like a finite graph, in the sense that the distribution of $\kappa(v)$ is non-degenerate – see Theorem 5.3. In order to evolve these two main results of this section, let us first properly define what we mean by "sufficiently many extra vertices".

Definition 7

Let $G = (V, E)$ be an infinite connected simple graph. It is said to contain a *neighbor-rich half-line*, if there exists a subgraph of G consisting of a half-line

$$H = \big(\{v_k,\ k \in \mathbb{N}\}, \{\langle v_k, v_{k+1}\rangle,\ k \in \mathbb{N}\}\big)$$

and distinct vertices $\{u_k,\ k \in \mathbb{N}\}$ from $V \setminus \{v_k,\ k \in \mathbb{N}\}$ such that there is an injective function $f : \mathbb{N} \to \mathbb{N}$ with the following two properties (cf. Figure 9):

(i) For all $k \in \mathbb{N}$: $\langle u_k, v_{f(k)}\rangle \in E$, i.e. the vertices u_k and $v_{f(k)}$ are neighbors in G.

(ii) The function f is growing slowly enough in the sense that $\sum_{k=1}^{\infty} \frac{1}{f(k)}$ diverges.
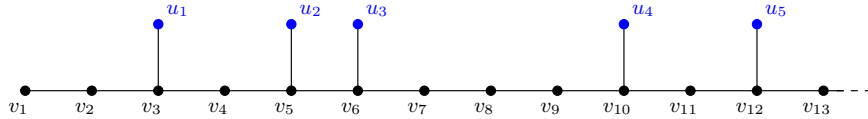

Figure 9: The beginning part of a neighbor-rich half-line.

Note that – by a renumbering of $\{u_k,\ k \in \mathbb{N}\}$ – we can always assume the function f to be (strictly) increasing. Furthermore, if G is connected and

contains a neighbor-rich half-line, we can choose any vertex $v \in V$ to be its beginning vertex: If v_l is the vertex with highest index at shortest distance to v in H, replace $(v_1, \ldots, v_l)$ by a shortest path from v to v_l in H. The altered half-line will still be neighbor-rich, since for all $M, N \in \mathbb{N}$ and f as above:

$$\sum_{k=1}^{\infty} \frac{1}{f(k)} = \infty \quad \Longleftrightarrow \quad \sum_{k=M}^{\infty} \frac{1}{f(k) + N} = \infty.$$

With this notion in hand, we can state and prove the following result:

Theorem 5.1
Consider an infinite (connected) graph $G = (V, E)$ and the initial water levels to be i.i.d. $\mathrm{unif}([0, 1])$. Let $v \in V$ be a fixed vertex of the graph. If G contains a neighbor-rich half-line, then $\kappa(v) = 1$ almost surely.

Before embarking on the proof of this theorem, we are going to show a standard auxiliary result which will be needed in the proof:

Lemma 5.2
For $\varepsilon > 0$, let $(Y_k)_{k \in \mathbb{N}}$ be an i.i.d. sequence having Bernoulli distribution with parameter ε. If the function $f : \mathbb{N} \to \mathbb{N}$ is strictly increasing and such that $\sum_{k=1}^{\infty} \frac{1}{f(k)}$ diverges, then

$$\sum_{k=1}^{\infty} \frac{Y_k}{f(k)} = \infty \quad \text{almost surely.}$$

PROOF: Let us define

$$X_n = \sum_{k=1}^{n} \frac{Y_k - \varepsilon}{f(k)} \quad \text{for all } n \in \mathbb{N}.$$

As the increments are independent and centered, this defines a martingale with respect to the natural filtration. Furthermore,

$$\mathbb{E}\left(X_n^2\right) = \sum_{k=1}^{n} \frac{\mathbb{E}\left(Y_k - \varepsilon\right)^2}{f(k)^2} = \left(\varepsilon - \varepsilon^2\right) \cdot \sum_{k=1}^{n} \frac{1}{f(k)^2} \leq \varepsilon \frac{\pi^2}{6}.$$

By the L^p-convergence theorem (see for instance Thm. 5.4.5 in [3]) there exists a random variable X such that X_n converges to X almost surely and in L^2. Having finite variance, X must be a.s. real-valued and due to

$$\sum_{k=1}^{n} \frac{Y_k}{f(k)} - X_n = \varepsilon \cdot \sum_{k=1}^{n} \frac{1}{f(k)},$$

the divergence of $\sum_{k=1}^{\infty} \frac{1}{f(k)}$ forces $\sum_{k=1}^{\infty} \frac{Y_k}{f(k)} = \infty$ almost surely. $\qquad \square$

PROOF OF THEOREM 5.1: Given a graph G with the properties stated and a vertex v, we can choose a neighbor-rich half-line H with $v = v_1$ and the set

30

of extra neighbors $\{u_n\}_{n\in\mathbb{N}}$ as described in and after Definition 7. The initial water levels at $\{u_n\}_{n\in\mathbb{N}}$ are i.i.d. unif$([0,1])$, of course.

Depending on the random initial profile, let us define the following SAD-procedure starting at v: Fix $\varepsilon, \delta > 0$ and let $\{N_l\}_{l\in\mathbb{N}}$ be the increasing (random) sequence of indices chosen such that the initial water level at u_{N_l} is at least $1-\varepsilon$ for all l. Then define the SAD-procedure – starting with $\xi_0(v) = 1$, $\xi_0(u) = 0$ for all $u \in V \setminus \{v\}$ – such that first all vertices along the line $(v_1, v_2, \ldots, v_{f(N_1)}, u_{N_1})$ exchange liquids sufficiently often to get

$$\xi_{k_1}(u_{N_1}) \geq \frac{1}{f(N_1) + 2} \quad \text{for some } k_1 > 0,$$

and never touch u_{N_1} again. Note that by Lemma 2.2, $\xi_k(u_{N_1})$ can be pushed as close to $\frac{1}{f(N_1)+1}$ as desired in this way. At time k_1, the joint amount of water in the glasses at $v_1, v_2, \ldots, v_{f(N_1)}$ equals $1 - \xi_{k_1}(u_{N_1})$ and we will repeat the same procedure along $(v_1, v_2, \ldots, v_{f(N_2)}, u_{N_2})$ to get

$$\xi_{k_2}(u_{N_2}) \geq \frac{1}{f(N_2) + 2} \cdot \big(1 - \xi_{k_1}(u_{N_1})\big) \quad \text{for some } k_2 > k_1$$

and iterate this.

After m iterations of this kind, the joint amount of water localized at vertices of the half-line H equals $1 - \sum_{l=1}^m \xi_{k_l}(u_{N_l})$, which using $1 - x \leq e^{-x}$ can be bounded from above as follows:

$$
\begin{aligned}
1 - \sum_{l=1}^m \xi_{k_l}(u_{N_l}) &\leq \prod_{l=1}^m \left(1 - \frac{1}{f(N_l) + 2}\right) \\
&\leq \exp\left(-\sum_{l=1}^m \frac{1}{f(N_l) + 2}\right).
\end{aligned}
\tag{13}
$$

Defining $Y_k := \mathbb{1}_{\{\eta_0(u_k) \geq 1-\varepsilon\}}$ for all $k \in \mathbb{N}$ we get $(Y_k)_{k\in\mathbb{N}}$ i.i.d. Ber(ε) and can rewrite the limit of the sum in the exponent as follows:

$$\sum_{l=1}^\infty \frac{1}{f(N_l) + 2} = \sum_{k=1}^\infty \frac{Y_k}{f(k) + 2}.$$

This allows us to conclude from Lemma 5.2 that the exponent in (13) tends a.s. to $-\infty$ as $m \to \infty$. Consequently, $m, T \in \mathbb{N}$ can be chosen large enough such that with probability $1 - \delta$ it holds that

$$\sum_{l=1}^m \xi_{k_l}(u_{N_l}) \geq 1 - \varepsilon \quad \text{and} \quad k_m \leq T.$$

Given this event, the move sequence corresponding to the SAD-procedure just described – adding no further updates after time k_m, i.e. $\mu_k = 0$ for $k > k_m$, if $k_m < T$ – then ensures (see Lemma 2.1) that

$$\eta_T(v) \geq \sum_{l=1}^m \xi_T(u_{N_l}) \eta_0(u_{N_l}) \geq (1 - \varepsilon)^2,$$

31

forcing $\kappa(v) \geq (1-\varepsilon)^2$ with probability at least $1-\delta$. Since $\delta > 0$ was arbitrary, this implies $\kappa(v) \geq (1-\varepsilon)^2$ a.s. and letting ε go to 0 then establishes the claim.

$\square$

Let us now take a look at how this result can be used to crystallize the outstanding leanness of the infinite line among all infinite quasi-transitive graphs. To this end, let us first repeat the definition of quasi-transitivity.

Definition 8

Let $G = (V, E)$ be a simple graph. A bijection $f : V \rightarrow V$ with the property that $\langle f(u), f(v) \rangle \in E$ if and only if $\langle u, v \rangle \in E$ is called a *graph automorphism*. G is said to be *(vertex-) transitive* if for any two vertices $u, v \in V$ there exists a graph automorphism f that maps u on v, i.e. $f(u) = v$.

If the vertex set V can be partitioned into finitely many classes such that for any two vertices u, v belonging to the same class there exists a graph automorphism that maps u on v, the graph G is called *quasi-transitive*.

Note that the notion of quasi-transitivity becomes meaningful only for infinite graphs as all finite graphs are quasi-transitive by definition.

Theorem 5.3

Consider an infinite (connected) quasi-transitive graph $G = (V, E)$ and the initial water levels to be i.i.d. $\text{unif}([0, 1])$. *Let $v \in V$ be a fixed vertex of the graph. If G is the line graph, that is $V = \mathbb{Z}$ and $E = \{\langle u, u + 1 \rangle,\ u \in \mathbb{Z}\}$, then $\kappa(v)$ depends on the initial profile. If G is not the line graph, then $\kappa(v) = 1$ almost surely.*

PROOF: Given i.i.d. $\text{unif}([0, 1])$ initial water levels, we can immediately conclude two things: If G is an infinite (connected) graph, the strong law of large numbers guarantees $\kappa(v) \geq \frac{1}{2}$ almost surely.

If G is the infinite line graph, there is a positive probability that the vertex v is what Häggström [4] calls two-sidedly ε-flat with respect to the initial profile (see La. 4.3 in [4]), i.e.

$$\frac{1}{m + n + 1} \sum_{u=v-m}^{v+n} \eta_0(u) \in \left[\tfrac{1}{2} - \varepsilon, \tfrac{1}{2} + \varepsilon\right] \quad \text{for all } m, n \in \mathbb{N}_0. \qquad (14)$$

La. 6.3 in [4] states that in this situation, the water level at v is bound to stay within the interval $[\frac{1}{2} - 6\varepsilon, \frac{1}{2} + 6\varepsilon]$ irrespectively of future updates. Together with the simple observation $\kappa(v) \geq \eta_0(v)$, it implies that $\kappa(v)$ is a random variable with non-degenerate distribution on $[\frac{1}{2}, 1]$.

In view of Theorem 5.1, to prove the second part, we only have to verify, that an infinite, connected, quasi-transitive graph that is not the line graph contains a neighbor-rich half-line. Since G is infinite (and by our general assumptions both connected and having finite maximal degree) a compactness argument guarantees the existence of a half-line H on the vertices $\{v_k,\ k \in \mathbb{N}\}$ such that $v_1 = v$ and the graph distance from v_k to v is $k - 1$ for all k.

Let us consider the function $d : V \to \mathbb{N}_0$, where $d(u)$ is the graph distance from the node u to a vertex of degree at least 3 being closest to it. Since G is quasi-transitive, connected and not the line graph, d is finite and can take on only finitely many values, which is why it has to be bounded, by $C \in \mathbb{N}$ say. Consequently, G can not contain stretches of more than $2C$ linked vertices of degree 2. For this reason, there must be a vertex among $v_3, \ldots, v_{2C+3}$, say $v_{f(1)}$, having a neighbor u_1 outside of H. In the same way, we can find a vertex u_2 outside H having a neighbor $v_{f(2)}$ among $v_{2C+6}, \ldots, v_{4C+6}$ and in general some u_k not part of H but linked to a vertex $v_{f(k)} \in \{v_k, \ k \in \mathbb{N}\}$ with

$$(k-1)(2C+3) + 3 \leq f(k) \leq k(2C+3) \quad \text{for all } k \in \mathbb{N}.$$

This choice makes sure that $v_{f(j)}$ and $v_{f(k)}$ are at graph distance at least 3 for $j \neq k$, which forces the set $\{u_k, \ k \in \mathbb{N}\}$ to consist of distinct vertices. Due to

$$\sum_{k=1}^{\infty} \frac{1}{f(k)} \geq \frac{1}{2C+3} \sum_{k=1}^{\infty} \frac{1}{k} = \infty,$$

H is a neighbor-rich half-line in the sense of Definition 7 as desired. $\qquad\square$

Remark

(a) Note that the essential property of the initial water levels, needed in the proof of Theorem 5.1, was independence. The argument can immediately be generalized to the situation where the initial water levels are independently (but not necessarily identically) distributed on $[0, C]$ and we have some weak form of uniformity, namely:

For every $\delta > 0$, there exists some $\varepsilon > 0$ such that for all $v \in V$:

$$\mathbb{P}\big(\eta_0(v) > C - \delta\big) \geq \varepsilon.$$

The sequence $Y_k := \mathbb{1}_{\{\eta_0(u_k) \geq C - \delta\}}$, $k \in \mathbb{N}$, similar to the one defined in the proof of Theorem 5.1 will no longer be i.i.d. $\text{Ber}(\varepsilon)$, but an appropriate coupling will ensure that

$$\sum_{k=1}^{\infty} \frac{Y_k}{k} \geq \sum_{k=1}^{\infty} \frac{Z_k}{k}$$

almost surely, where $(Z_k)_{k \in \mathbb{N}}$ is an i.i.d. sequence of $\text{Ber}(\varepsilon)$ random variables. Accordingly, we get $\kappa(v) = C$ a.s. even in this generalized setting.

(b) As alluded to in the introduction, the statement of Theorem 5.3 can be interpreted in the following way: When it comes to the qualitative behavior of $\kappa(v)$ for a fixed vertex v in the graph, the radical change does not happen between finite and infinite graphs but rather between the line graph $\mathbb{Z}$ and all other quasi-transitive infinite graphs, which is why the results for the Deffuant model on $\mathbb{Z}$ can not immediately be transferred to higher-dimensional grids – as discussed in the introduction of Sect. 3 in [5].

(c) Finally, it is worth emphasizing that Theorem 5.3 does not capture the full statement of Theorem 5.1: If we take the infinite line graph $\mathbb{Z}$ and add an extra neighbor to every node that corresponds to a prime number, the only quasi-transitive subgraph contained is the line graph itself. However, since it contains a neighbor-rich half-line, Theorem 5.1 states that $\kappa(v) = 1$ for i.i.d. unif($[0, 1]$) initial water levels and any target vertex v.

References

[1] COOK, S.A., The Complexity of Theorem-Proving Procedures, *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, pp. 151-158, 1971.

[2] COX, J.T., GANDOLFI, A., GRIFFIN, P.S. and KESTEN, H., Greedy lattice animals I: upper bounds, *Annals of Applied Probability*, Vol. 3 (4), pp. 1151-1169, 1993.

[3] DURRETT, R., *"Probability: Theory and Examples (4th edition)"*, Cambridge University Press, 2010.

[4] HÄGGSTRÖM, O., A pairwise averaging procedure with application to consensus formation in the Deffuant model, *Acta Applicandae Mathematicae*, Vol. 119 (1), pp. 185-201, 2012.

[5] HÄGGSTRÖM, O. and HIRSCHER, T., Further results on consensus formation in the Deffuant model, *Electronic Journal of Probability*, Vol. 19, 2014.

[6] LEE, S., An inequality for greedy lattice animals, *Annals of Applied Probability*, Vol. 3 (4), pp. 1170-1188, 1993.

[7] SHANG, Y., Deffuant model with general opinion distributions: First impression and critical confidence bound, *Complexity*, Vol. 19 (2), pp. 38-49, 2013.

OLLE HÄGGSTRÖM
DEPARTMENT OF MATHEMATICAL SCIENCES,
CHALMERS UNIVERSITY OF TECHNOLOGY,
412 96 GOTHENBURG, SWEDEN.
olleh@chalmers.se

TIMO HIRSCHER
DEPARTMENT OF MATHEMATICAL SCIENCES,
CHALMERS UNIVERSITY OF TECHNOLOGY,
412 96 GOTHENBURG, SWEDEN.
hirscher@chalmers.se